Tampere University

Milla Laukkanen

CONTRACTIONS *GONNA* AND *WANNA* IN INTERNET LANGUAGE:

A corpus study

Faculty of Information Technology and Communication Sciences
Master's Thesis
April 2020

ABSTRACT

Milla Laukkanen: Contractions *gonna* and *wanna* in internet language: A corpus study
Master's Thesis
Tampere University
Master's Programme in Languages
April 2020

This thesis examines the use of the contractions *wanna* and *gonna* in written internet language based on corpus data. These contractions are typically used only in spoken language due to their informality, and they rarely occur in texts that use formal language. The aim of the thesis is to discover how the use of the contractions is distributed across the different registers of internet language. The thesis also intends to determine how the contractions are used in these registers in comparison to their full forms *want to* and *going to*.

The data is gathered from the iWeb corpus, which consists of a broad collection of web pages. The data is categorized into web registers using the 2015 hierarchical framework of Biber et al, which is introduced in the theory section of the thesis. This chapter also introduces background literature related to internet language and explains its relationship to spoken and written language. In addition, the theory section presents some previous research that has been done on the contractions.

The analyzed data reveals that the contractions *wanna* and *gonna* are the most frequent in contexts that resemble conversation. The contractions are most often used in the interactive discussion web register, and more specifically in its sub-registers of discussions forums and reader comments. The use of the contractions is largely limited to these interactive situations, although they also occur to some extent in registers that are directly connected to speech. Such situations are song lyrics, transcripts of video and audio, and TV and movie scripts. The contractions were also compared to their full forms *want to* and *going to*, and the results indicate that the use of the full forms is distributed more evenly across the different registers. These results were expected, as the longer word forms also occur in written texts outside of the internet and their use is not generally limited to spoken language.

The findings indicate that the contractions typical to spoken language occur in written online texts as well. The writing style of discussion forums and reader comments prefers the contractions over the full forms, and thus closely resembles speech. This can be explained by the similarities between these registers and face-to-face conversations, although the communication that takes place over the web has some limitations. Interactivity and the number of participants are also essential to these registers. Additionally, registers that are directly linked to speech prefer to use the contractions over the full forms as well. Other registers use mainly the full forms, and consequently their language resembles the formal style of traditional written texts.

Keywords: internet language, register, contraction, corpus linguistics

TIIVISTELMÄ

Milla Laukkanen: Contractions *gonna* and *wanna* in internet language: A corpus study
Pro gradu -tutkielma
Tampereen yliopisto
Kielten maisteriohjelma
Huhtikuu 2020

Tutkielma tarkastelee englanninkielisten lyhenteiden *wanna* ja *gonna* käyttöä internetkielessä korpuksen avulla. Käsiteltäviä lyhenteitä käytetään tyypillisesti englannin puhekielessä niiden epämuodollisen tyylin vuoksi, eikä niitä esiinny juurikaan virallista kieltä käyttävissä teksteissä. Tutkielman tarkoituksena on tutkia lyhenteiden käytön jakautumista internetkielen eri tekstilajien välillä. Tarkoituksena on myös selvittää, miten lyhenteiden käyttö internetin tekstilajeissa eroaa niiden pidemmistä muodoista *want to* ja *going to*.

Tutkielman aineisto on kerätty iWeb korpuksen laajasta verkkosivujen kokoelmasta. Aineisto jaetaan eri tekstilajeihin käyttäen apuna Biber et al. vuoden 2015 hierarkkista viitekehystä, joka esitellään tutkielman teoriaosuudessa. Teoriaosuus esittelee myös aiheeseen liittyvää taustakirjallisuutta internetkielen osalta, sekä selittää kyseisen kielen suhteita puheeseen ja kirjoitettuun tekstiin. Osiossa käsitellään myös lyhenteistä *wanna* ja *gonna* tehtyjä aikaisempia tutkimuksia.

Analysoidusta aineistosta selviää, että lyhenteitä *wanna* ja *gonna* esiintyy eniten keskustelua muistuttavissa konteksteissa. Lyhenteitä käytetään eniten interaktiivisen keskustelun tekstilajissa, ja tarkemmin sen alalajeissa keskustelufoorumeilla sekä sivujen kommenttiosioissa. Lyhenteiden käyttö on keskittynyt lähinnä näihin interaktiivisiin keskusteluihin, mutta niitä esiintyy myös jonkin verran puheeseen suoraan linkittyvissä tilanteissa. Tällaisia tekstilajeja ovat laulujen lyriikat, äänitteiden ja videoiden litteraatiot sekä elokuvien ja tv-sarjojen käsikirjoitukset. Lyhenteitä vertailtiin myös niiden pidempiin muotoihin *want to* ja *going to*, joiden käyttö oli jakautunut tasaisemmin eri tekstilajien välillä. Nämä tulokset olivat odotettuja, sillä pidempiä muotoja esiintyy kirjoitetuissa teksteissä myös internetin ulkopuolella eikä niiden käyttö ole lyhenteiden tavoin rajautunut puhekieleen.

Näiden tulosten perusteella on mahdollista todeta, että tyypillisesti vain puheessa esiintyneitä lyhenteitä *wanna* ja *gonna* käytetään myös kirjoitetussa muodossa internetissä. Internetkielen tekstilajeista keskustelufoorumeiden ja kommenttiosioiden kirjoitustyyli suosii eniten epämuodollisia lyhenteitä ja muistuttaa näin eniten puhetta. Tämä voidaan selittää sillä, että kyseiset tekstilajit muistuttavat tyypillistä kasvotusten tapahtuvaa keskustelua, vaikkakin internetin kautta tapahtuvissa keskusteluissa on rajoitteita. Kyseisiin tekstilajeihin liittyy olennaisina tekijöinä myös niiden interaktiivisuus ja keskustelijoiden lukumäärä. Myös puheeseen suoraan linkittyvät tekstilajit suosivat lyhenteitä pidempien muotojen sijaan. Muut tekstilajit puolestaan käyttävät lähinnä pidempiä muotoja, ja nämä sivut muistuttavat enemmän muodollista kieltä suosivia kirjoitettuja tekstejä.

Avainsanat: internetkieli, tekstilaji, lyhenne, korpuslingvistiikka

Table of Contents

1. Introduction 1
2. Theoretical background 5
   2.1. Internet language 5
   2.2. Written and spoken aspects of internet language 8
   2.3. Web registers 10
   2.4. Previous research on *wanna, gonna* and their full forms 13
3. Data and methods 17
4. Results 23
   4.1. *Wanna* 23
   4.2. *Want to* 26
   4.3. *Gonna* 28
   4.4. *Going to* 31
5. Discussion 34
   5.1. Comparisons of *wanna* and *want to* 34
   5.2. Comparisons of *gonna* and *going to* 37
   5.3. Hybrids 38
   5.4. Patterns and variation 39
   5.5. General discussion 49
6. Conclusion 55

References 58

1. Introduction

Internet is a place for humans to interact with each other through text, regardless of their physical location. This concept is known as computer-mediated communication, which typically refers to text-based communication between the participants (Herring 1996: 1). Some examples of computer-mediated communication are discussion forums and blogs, which do not exist in the offline world and can result in innovative uses of literate forms to compensate for the lack of face-to-face interaction (Barton & Lee 2013: 16). Internet allows the study of how language can evolve to support these new communicative situations (Baron 2008: 4). Consequently, analyzing the characteristics of internet language helps us to understand language overall better (McCulloch 2019: 4).

Internet language is special in the way it combines features of spoken and written language in its text. Writing is a crucial part of communication on the internet, but the linguistic forms that appear in online writing can be informal and closely resemble spoken language. Herring defines the language of computer-mediated communication as typed text that is written, but also used in fast exchanges that are informal and mirror spoken conversations (1996: 3). Furthermore, McCulloch explains language online as a combination of writing and informality that is changing the way communication can take place. There are norms that exist when writing texts for newspapers or other formal platforms, and these same norms are not present in the informal contexts of forums and comment sections (McCulloch 2019: 7–8).

Defining some internet texts as informal therefore suggests that informal expressions can be expected to occur in certain online contexts. This indicates that contractions typical to spoken language, e.g. *wanna* and *gonna*, can appear in internet language as well. Nevertheless, computer-mediated communication has several areas where language is used, and they all do not fall into the category of informal writing. Internet has both formal and

informal spaces where communication takes place. There are numerous categories of web pages that use language differently, and this diversity in internet language makes it a topic worth studying (Crystal 2011: 18). Classifying internet language into registers leads to a better understanding of its use and supports the study of its linguistic features (Biber et al. 2015b: 1817).

This thesis is a corpus-based study of the use of informal expressions *gonna* and *wanna* in internet language. *Gonna* and *wanna* have the same meanings as their full forms *going to* and *want to*, and they have been referred to as their contractions (Bolinger 1981, Hudson 2006), phonetically reduced structures (Machová 2015), reduced non-standard spellings (Leech et al. 2009), and variant forms (Berglund 2005). In this thesis the forms are referred to as contractions and informal word forms, as the study focuses on how phrases with corresponding meanings are used in different online contexts and how informal and contracted forms differ from their full and more formal forms.

Contracted forms are characteristic of speech, especially in informal contexts, and these shorter word forms are not typically used in writing (Crystal 2011: 26). However, contractions do occur in internet language, which raises questions about their usage patterns and the reasons why they are used in a written form at all. Therefore, the aim of this thesis is to examine how common *wanna* and *gonna* are in web registers in comparison to their full forms, and what patterns can be found in their use. I will approach the topic through two research questions:

1. Which web registers show the highest number of occurrences of *wanna* and *gonna*, and why?
2. How does the distribution of the contracted forms differ from their full forms across the web registers?

The hypothesis of the study is that *wanna* and *gonna* are used more in informal online contexts that imitate conversation, such as discussion forums and comment sections, and less in formal contexts that do not resemble spontaneous conversation. As a result, it is possible to define certain web registers as more speech-like than others. Of course, the contractions are also expected to occur often in web registers that are directly linked to spoken language, such as interviews and transcripts. The contracted forms are the focus of the study, but they are also compared to their full forms for a deeper understanding of their usage patterns. These comparisons will provide information about which forms are preferred in each web register.

Examining only the occurrences of the contractions would be insubstantial, as their relation to the rest of the language use in the registers could not be analyzed. Studying the distribution of the full forms therefore supports the analysis of the contractions. The full forms *want to* and *going to* are expected to occur more consistently across all the web registers. This is because the use of these forms is traditionally not limited to spoken contexts, unlike the contractions. Therefore, the full word forms can be expected to be commonly used on a variety of web pages.

Studies on internet language have often focused on the synchronous communication of chats and social media, perhaps due to them differing greatly in mode from traditional spoken and written communication. This study looks at internet language from another perspective, regarding searchable web. Searchable web is defined as the content discoverable through a search engine, rather than the entire web that cannot be approached systematically (Biber et al. 2015b: 1830). This part of the internet does not consist of only short clauses and swiftly written strings of texts, as is often the case with chats and social media postings. Instead, the participants tend to write longer sections of text. Nonetheless, these texts are not necessarily any more edited than those that appear on social media or in instant

messages, which can result in high frequencies of informal word forms. It will therefore be interesting to see how the use of contractions is distributed in this part of the web, rather than in synchronous communication.

Both of the *wanna* and *gonna* contractions have been studied previously from a variety of perspectives (Verplaetse 2003, Berglund 2005, Hudson 2006, Leech et al. 2009, Lorenz 2013, Machová 2015), but their use across different web registers has not been focused on. This thesis fills a research gap on how exactly these typically spoken expressions are used in the written mode across the searchable web. Comparing the contractions to their full forms word forms will also indicate which forms are preferred in certain registers of internet language. This can lead to the categorization of some web registers as more informal than others. Furthermore, some web registers can be found to resemble spoken conversation more than others, and as a result the different areas of internet language can be defined in more detail.

2. Theoretical background

There are several different areas of background theory that are used as the base of the thesis and explored in the following sections. To begin, section 2.1. will explain what internet language is and how language is used online. Section 2.2. will approach the relationship of written and spoken language online to account for how these two aspects of language are connected. Section 2.3. will provide background information about what web registers are, and about the theory used to define them. The advantages and difficulties of using this theory will also be considered. Finally, section 2.4. will present some of the previous research that has been done on the contracted forms *wanna* and *gonna* in relation to their full forms *want to* and *going to*.

2.1. Internet language

Internet language is computer-mediated communication that occurs between humans that interact through computers, and it is part of internet linguistics. David Crystal is a notable name on this field, and he has emphasized that computer-mediated communication is too broad as a term to describe the study of language online, as this category also includes videos, pictures, and other things that do not use only language. As a result, Crystal defines the study of language online as internet linguistics, to focus on the language aspect of communication (Crystal 2011: 11–2).

Internet language differs from regular conversations in a variety of ways. Firstly, the relationship between the sender and receiver is different, as anonymity plays a role in this form of communication by creating different expectations of how the conversation will develop in contrast to a discussion in a familiar face-to-face setting (Crystal 2011: 15). Additionally, the reader of an online text differs from the reader of a text that is printed on actual paper

because the setting is not as governed by traditional behavioral norms, and the text can be ambiguous and even offensive (Crystal 2011: 16). Herring notes several features of internet language that are unique to this genre, such as emoticons and other graphic features and acronyms (1996: 3). Another defining characteristic of language online is how its users adapt to the medium and innovate new ways of interaction to compensate for the lack of feedback and clear turns between the conversation participants. There are also advantages to online communication in written form: the participants use this platform to play with language, they have the ability to participate in multiple conversations simultaneously, and due to saved textual records they can fluently continue conversations after time has passed (Herring 1999: 10).

Rules that guide the use of language online are beginning to form, defining the web as a place where informal writing is strongly present (McCulloch 2019). McCulloch refers to standard language and spelling conventions as a collective agreement that can change among language users (2019: 46). Language users can collectively choose to use specific features online, which may result in spoken language forms appearing often in written texts. Informal language is increasingly being written down online in different contexts, which results in variation across all the different web registers. Each internet user decides which word forms they use online, and all these personal decisions result in collective language habits, also known as norms (McCulloch 2019: 56–7).

Norms are social cues that define how the participants act and react to things. These norms are linked to different social relationships, and the members of a certain social group behave in a way that is appropriate for the situation. Concerning communication online, different communicative situations can have their own norms that define what kind of language use is appropriate, and these norms may differ greatly from other situations. Specific online settings can allow or even encourage non-standard spelling, which would not be appropriate in

formal written texts (Hård af Segerstad 2003: 19). Thus, language online can resemble conversations, as its norms are similarly fluid and they can change based on the participants and their needs (McCulloch 2019: 198–9).

McCulloch mentions that language is made more efficient online by shortening words and phrases, for example by using acronyms and creating new ones. This is especially noticeable in informal writing, where the number of typed letters decreases (2019: 11). Shortening of phrases in these contexts can therefore be linked to word forms such as *wanna* and *gonna* that are contracted from longer phrases in speech. When these forms occur in written contexts, it is likely that their user wishes to communicate efficiently and in a way that imitates conversation. Previously it has been assumed that the use of shortened forms online would be purely based on the laziness of the language users, but instances of teens using a mixture of formal and informal writing styles online suggests that the use of contracted forms online is a choice made by the language users. This mixture of writing styles indicates that internet language is not a transcript of speech or an attempt to write formally, but instead something else entirely. Accordingly, McCulloch defines internet language as its own genre with specific goals (2019: 60).

However, this definition assumes that all of internet language is one genre that follows the same rules, while in reality there is variation within the entity of internet language. Communicative functions can vary immensely across different web pages, from factual content to informal small talk (Crystal 2011: 29). Internet language has different styles and categories that it can be divided into, similarly to other forms of language. These categories are partly based on the technology that is available, and they are also defined by human factors such as the communicative purpose of the event (Herring 1996: 3–4). New technology may arise that affects the communicative event, or the technology may be the same and the social context evolves (McCulloch 2019: 235). Internet language therefore has multiple modes that have their

own restrictions and benefits, and as a result communication on this platform cannot be treated as a singular unity (Hård af Segerstad 2003: 36).

When communication online is discussed, a distinction should be made between the two types of communication there are, namely asynchronous and synchronous communication. Asynchronous communication is not dependent on time or place, and the participants can take part in the discussion freely. This resembles traditional written texts that are not dependent on temporal or spatial factors. Some examples of this type of communication are emails and discussion forums. Contrastively, synchronous communication requires the participants to be online at the same time so that the discussion can be successful. Synchronous discussions have been characterized as interactive, although interactivity is not limited only to this type of online communication. An example of synchronous communication is web chat, where the participants are required to be online simultaneously (Hård af Segerstad 2003: 4). Division between these two types of communication is not relevant for all genres of internet language, but it should be taken into consideration whenever the different areas of computer-mediated communication are discussed.

2.2. Written and spoken aspects of internet language

As the internet has grown, so has the role of language evolved to reflect the need for fluent communication online through written text. Thus, a question has risen about whether this way of using language is closer to speech or writing (Baron 2008: 28). Herring has defined computer-mediated communication as typed language that resembles writing, but also borrows informal features from spoken language and conversations (1996: 3). Collot & Belmore further elaborate on this relationship of spoken and written language in their 1996 article regarding electronic language. They describe internet language as neither spoken nor written, as the users often follow the fast-paced interactional style of spoken conversations, but the communication

does not take place face-to-face or through actual speech. In addition, internet language users do not always plan out or edit their text in a way that is typical to writers (Collot & Belmore 1996: 14).

Yates (1996) studied computer conferencing data from the aspects of written and spoken language and found features of both areas of discourse. Yates defines speech as a form of communication that is produced rapidly between the participants and heard in the same way. Contrastively, writing is defined as static, due to it being produced at a pace that the writer chooses and consumed by the other receiving participants when they decide to do so. According to Yates, these differences result in likely variation in the language that is used in speech and writing (1996: 33). The differences between the two forms of communication also influence internet language, as it adopts features of both speech and writing. Language online uses vocabulary that is typical to writing but passes on information in ways that resemble speech (Yates 1996: 35). However, internet language also resembles speech in its use of pronouns and modal auxiliaries. Language online is therefore a mixture of these two and it is affected by a variety of social structural and social situational factors that ultimately define it (Yates 1996: 46).

Informality is also a factor that should be considered when discussing the spoken and written aspects of internet language. Outside the internet, personal notes and letters are written in a casual manner. Still, writing is viewed more often as formal while speech is considered informal. Students are taught to not use contractions in school, and language users are not encouraged to write informally (Baron 2008: 46). Internet language is defined by many as a combination of speech and writing, and the persistence of some formal text features online can be explained by habit strength. Formal writing style is expected throughout one's education, and therefore it is difficult to ignore the rules even if internet would allow it (Baron 2008: 70). Nevertheless, certain features of spoken language do appear in writing online,

perhaps because the conversational part of internet communication is guided by speech conventions (Crystal 2011: 29). Despite this, there is uncertainty about how to exactly define internet language today. People refer to emails as conversations, but one still "writes" an email instead of speaking it and "reads" web pages (Crystal 2011: 25). Thus, defining internet language as one or the other is undoubtedly complex.

Internet language has therefore been defined in a variety of ways and speech and writing are both established to be a crucial part of it. Nevertheless, none of these definitions of internet language make a clear division of the different registers that exist within this medium and what differences and similarities they have. Some parts of internet language resemble spoken language, while others are more alike to written language (Collot & Belmore 1996: 18). Hård af Segerstad has argued that it is not clear-cut what is considered spoken or written language, as people adapt to the situation and use language accordingly (2003: 3). Overlapping between spoken and written language takes place online as well, which raises questions about how exactly this externalizes across the different registers of internet language.

2.3. Web registers

This section will explain how web pages can be divided into registers and sub-registers based on the theoretical framework of Biber et al. (2015a & 2015b). This theory will be used as the base of the analysis in this thesis, but it is also central to the study of internet language today. According to Crystal, internet is in a constant state of transition, with the people expressing themselves in different ways in various situations (2006: 16). The lack of universal written rules in internet language is distinctive, which means that the people who use language there can essentially create the rules and conventions themselves (Crystal 2006: 16). Web pages can be analyzed based on the situation where the language is used, and this defines which register the specific web page belongs to. Registers in this context are defined as language varieties that

are associated with a certain situation of use and a specific communicative purpose (Biber et al. 2015b: 1820).

Web register is therefore the style of the language that follows the norms of the setting where it occurs. These online register categories include the textual distinctions that exist in standard written corpora, as well as ones that are typical to unpublished online texts (Biber et al. 2015a: 14). Some examples of registers that fall under the first definition are novels, research articles and news reports. In contrast, the second category includes blogs, discussion forums, and other web pages that are not published and do not have an equivalent paper form outside the internet. Internet includes registers that fit into both categories, as well as ones that fall somewhere in between. All these registers are defined by their situational characteristics, similarly to the registers outside the internet.

The defining situational characteristics of web registers include the participants of the action, its interactivity, the communicative purpose of the situation, and the chosen topic (Biber et al. 2015b: 1818). Web pages typically have a limited amount of external indication about which register category they can be defined as, compared to other written and published texts (Biber et al. 2015a: 13). A variety of texts that exist on actual paper have a sufficient number of external features that simplify their classification into register categories, e.g. sport reports that are published in the sports section of a newspaper. Some web pages can be easily categorized as well, but their external factors do not define them as clearly into register categories (Biber et al. 2015a:14).

To simplify the categorization of web pages, Biber et al. have built a framework that follows the shape of a hierarchical decision tree. This framework results in the web pages being divided first into general register categories, and later into more specific sub-registers based on the first choice of category. At the top level of the framework the internet text can be divided in two ways, based on whether it has been produced as speech or in a written mode. A

text that has been produced in the spoken mode can then be categorized accordingly, for example as *speech* or *interview*. (Biber et al. 2015a: 19).

A written text is further divided between an *interactive discussion* that involves multiple participants, and a non-interactive text. Furthermore, a non-interactive text is defined based on its communicative purpose, resulting in the remaining six general categories. These communicative purposes include narrating or reporting events, describing or explaining information, expressing opinion, describing or explaining facts with intent to persuade, explaining how-to or instructions, and expressing oneself through lyrics (Biber et al. 2015a: 19–20). After the text has been categorized into one of these web register categories based on its characteristics, it is further divided into sub-registers under each general category. One example is the sub-category of *travel blog*, which falls under a general category that has the characteristics of being written, noninteractive, and narrative (Biber et al. 2015a: 20).

Biber et al. used multiple participants in their analysis to rate the web pages and divide them into the web register categories. Furthermore, after comparing the answers they found different levels of agreement between the participants when dividing the web pages into the categories, particularly into the sub-registers. The answers did not always correspond with those of the other participants. This means that using the framework as a base of categorizing web pages is not fully reliable, as each person may categorize a web page slightly differently. However, the web pages were largely classified in a unified way into the general web registers (Biber et al. 2015a: 20). This is likely because the framework is more reliable as it uses a hierarchical approach that is based on the situational characteristics and the communicative purpose of the web pages. The sub-registers do not have a similar detailed framework to support it, which suggests that the general register categories provide more reliable results (Biber et al. 2015a: 22). Nevertheless, Biber et al. found that reliability in sub-register categorization was higher when the category was chosen from a short list of related register categories, instead of

picking it from a list that included all the sub-registers (2015b: 1822). The difference in the categorization is possibly also due to the divided nature of certain web pages that were defined as *hybrids*. These are registers that share the characteristics of two or more registers and are therefore categorized as their combination (Biber et al. 2015a: 20). Hybrid registers do not fit into the framework as neatly as regular registers, which may result in inconsistency in their categorization.

The different levels of agreement in categorizing the web registers shows how the communicative purposes of the web pages help to reliably categorize them but can also result in hybrid registers that complicate the categorization. Comparing these results to the categorization of published written texts outside of the internet, there are clear differences. Published written texts typically follow strict register definitions, while categorizing web pages is not as straightforward and their classification requires a more detailed analysis of all their defining characteristics (Biber et al. 2015a: 40). A hierarchical register framework uses more than one level of abstraction to define the web pages, which aims to help reduce these issues in their classification (Biber et al. 2015b: 1829).

The framework is therefore a necessary aid when the aim is to study the different areas of internet language and variety within this entity. Dividing each web page into both the main register category as well as its sub-register provides a detailed definition of each example. In this thesis the registers are used as a base for the analysis of the use of contractions. This framework is altered slightly to fit the purpose of this study, and its use will be discussed further in chapter 3.

2.4. Previous research on *wanna*, *gonna* and their full forms

The previous sections have explained the theoretical background of internet language, its relationship with spoken and written language, and how it can be divided into web registers.

Next the focus will be on how the contractions *wanna* and *gonna* have been studied previously. Often the words have been studied alongside their formal forms, for example by Verplaetse who focused on the modal status of *wanna* and *want to* and their relation to the expression of volition (2003). There have also been debates about whether *wanna* is actually considered a word, which is mentioned for example in Hudson's 2006 article. Hudson decides to treat *wanna* as two words on a syntactical level (i.e. *want to*), that are presented in a single word form. Hudson suggests that the last vowel of *wanna* can alternate in speakers' language, further explaining how this one word actually includes the meanings of the two words of the full form (2006: 604).

The words have also been studied in relation to the speaker's age. Leech et al. note that the shortened forms are more common in the speech of young speakers and less so among older speakers (2009: 106). Differences between spoken and written corpora have also been examined, with results indicating that the contractions *gonna* and *wanna* are more common in the spoken corpora and less frequent in the written corpora where the full forms *going to* and *want to* are in the majority (Leech et al. 2009: 106).

Another study that approaches this topic is by Berglund (2005), who studied future references in corpora. She also looked at the usage of *gonna* and *going to* between written and spoken corpora, investigating whether either variant is used more in certain text categories. Berglund found that *gonna* and *going to* are both used in similar lexical and syntactic contexts, and they are used more often in spoken texts than in written ones (2005: 165–6). Berglund mentions that *gonna* is often referred to as a spoken informal variant of *going to*, and confirms this by her findings that *gonna* is almost exclusive to spoken corpora, while occurrences of *going to* vary across different spoken corpora (2005: 75). Comparisons were made between *gonna* and *going to* in the written part of the BNC, and the results indicate that the contraction was the most frequent in texts that consist of quoted speech as well as in other

contexts that resemble speech. This further establishes the contraction as a spoken variant in written texts (Berglund 2005: 84). *Gonna* also often co-occurred with slang words and non-standard constructions in written texts (Berglund 2005: 160).

Relating to the formality of these two forms, the spoken part of the BNC shows instances of *gonna* appearing more in informal settings, while *going to* is more frequent in formal text types (Berglund 2005: 101). Berglund also looked at the frequency of expressions of future in different text categories in the written part of BNC. The results show that while *gonna* was overall rare in the corpus, it had similar occurrence patterns to *going to*, as both forms were the most frequent in imaginative texts (Berglund 2005: 100). Berglund's study therefore provides a considerable amount of information about the use of *gonna* in different text categories, although the study does not focus on the online usage of the contraction.

Berglund and Leech et al. both mention that transcription practices complicate the analysis of these two forms in spoken corpora due to the differences in how the transcriber may have heard the use of the specific word forms (Berglund 2005: 75–6, Leech et al. 2009: 105). An active choice has therefore always been made about how to transcribe this specific word. However, in the case of my thesis this is not an issue, as I am using a corpus that is based on written texts. There is no need for transcriptions, and they do not therefore influence the results of the study. Another limitation of Berglund's study was the lack of instances of *gonna* in the corpora, which she mentions throughout her book (2005). This made the observations about its use questionable, and further research about its patterns of usage necessary.

One of the more recent pieces of research is Machová's 2015 article about the grammaticalization of the structures *gotta*, *gonna*, *wanna* and *better*. Machová uses corpora and web forums in her study to evaluate how grammaticalized these structures are, and her findings indicate that the ability of these structures to express several meanings at the same time are linked to their multiple functionalities (2015). Lorenz (2013) has also approached the

changing role of *wanna*, by studying the changing state of *gonna*, *gotta* and *wanna* towards independent lexical items. Lorenz came to the conclusion that the contracted form *wanna* is actually more common in spoken American English than the full form *want to*, and that there are clear changes regarding word forms taking place inside the minds of language users (2013: 34).

To conclude this section, the informal forms *wanna* and *gonna* have been studied from several different perspectives, but the focus has not yet been strictly on their written usage online and how they are distributed across different web registers. By studying the contractions online, this thesis aims to fill the research gap on this topic and find out what are the characteristics of the web registers that use them the most. Consequently, this will help to define the relationship of spoken and written language across the web registers. This will also provide a base for conducting possible further research on the use of contractions online.

3. Data and methods

The data that is used in this study was gathered from the iWeb corpus, where the content has been collected from a variety of websites. The corpus consists of 14 billion words from 94,391 websites that have been chosen in a systematic way from the searchable web. One criterion for these websites is that the content has been created mainly by English speakers, which makes the corpus a good source of data for the study of the English language.

One advantage of studying a language through data that has been gathered from the internet is that the material is already in a digital form, which makes it quick to access and analyze. The material will also not get distorted by the researcher's presence, as the text has been created unselfconsciously beforehand. It is also open-access data that is gathered after it has been shared with an audience. Furthermore, the data provides insight into how language works in an everyday environment where conversations between humans occur in a written form (McCulloch 2019: 5). Internet language provides a substantial number of examples of informal written language that has been previously limited to letters, diaries and other typically handwritten texts (McCulloch 2019: 4). Internet therefore provides a strong base for gathering comprehensive data of the different aspects of language in a variety of contexts.

The iWeb corpus was chosen as the source of the data because it provides a large amount of material from the internet that is both relevant for the study and easily searchable. The corpus has excluded websites that do not include enough analyzable text to be usable for the study of language. These websites consist mainly of pictures, or they have in other ways a limited amount of text. Thus, the texts that are included in the corpus represent how English is used in the textual part of the internet, and this data can be further analyzed from additional aspects.

The compilers of the corpus have gathered its data from the searchable web, which means that it includes formal texts, as well as examples of informal written language.

There is also documentation of both informal and formal speech events. Data gathered from this corpus therefore includes language use in several contexts that are defined by different situational characteristics. Consequently, the data provides a good number of examples from the different web register categories. As the compilers of the iWeb corpus have gathered its data from the open and searchable web, it does not include instances of synchronous communication, such as web chats and instant messages. This makes it especially suitable to be categorized into the web registers defined by Biber et al, as these types of communication situations are not included in the framework due to its focus on the searchable web. The synchronous forms of online communication have already been studied extensively, and this thesis concentrates on the other parts of the internet.

For this study, a random sample of 300 examples of each *wanna, gonna, want to* and *going to* have been gathered from the iWeb corpus, resulting in a total of 1,200 examples. Previous studies have struggled to find enough occurrences of the contractions to find any reliable patterns of their use in written texts, but this is not an issue here due to the size of the corpus. The contractions are both extensively displayed in the corpus when followed by a verb, *wanna* with 111,227 occurrences and *gonna* with 262,204 occurrences. While the numbers of both word forms are high, the reason I have limited my sample size to 300 per expression is due to the amount of work it takes to categorize each example. I have also saved additional examples of all phrases in case multiple web pages are inaccessible and thus cannot be analyzed.

The contractions *wanna* and *gonna* were chosen because they both provide a good number of instances in the iWeb corpus. It should be noted that only those usages of *want to* and *going to* were included in the study that could be used interchangeably with their informal counterparts. In the case of *going to*, this means excluding the cases where the phrase is not followed by a verb, because the contraction *gonna* is only used in such a context as it

refers to a future action. Similar restrictions were applied to *wanna*, and only those instances of it were included where it occurred before a verb. Table 1 below illustrates some of these excluded instances. These examples use *wanna* and *going to* before nouns, and they are therefore not comparable with the forms *want to* and *gonna* that occur before verbs.

When i grow up, i **wanna** map like this!
They **wanna** spectacle they want energy they want entertainment.
It sounds like you are **going to** college.
She barely has time to think after **going to** work and then cooking me three dinners.

Table 1. Excluded instances of *wanna* and *going to*.

Wanna can be used as a contraction of both *want to* and *want a*, but only the meaning of *want to* is studied in this thesis due to it being more common and therefore more comparable to the *gonna* contraction. Examples of the use of each word form in the data are visible in Table 2. As the table shows, each of the examples included in this study uses the word form before a verb.

Word form	**Context**
Wanna	All of this is still brand new to me and I don't **wanna** buy something and then realize it was a big waste of credits.
Gonna	Does he seem like someone who is **gonna** do something clever or new online?
Want to	The small taste of the city that we got on our Jordan trip made me **want to** spend a lot more time there.
Going to	And of course, everyone wants to know, how much is it **going to** cost me at the end of the day?

Table 2. Examples of the context of each word form included in this study.

The data is analyzed by using the 2015 register framework by Biber, Egbert & Davies, which was introduced in section 2.3. The importance of a hierarchical framework that classifies web pages into registers has been emphasized as a requirement to better understand the language of the internet (Biber et al. 2015b: 1817). The framework was explained in detail in the theory section, and Table 3 demonstrates how exactly the situational characteristics of a web page define its register.

Mode	Originally written							Originally spoken
Participants	Single authors or co-authors						Multiple participants	
Purpose	To narrate events	To describe information	To express opinion	To use facts to persuade	To explain instructions	To express lyrically		
Register	Narrative	Informational description/explanation	Opinion	Informational persuasion	How-to/ instructional	Lyrical	Interactive discussion	Spoken
Sub-registers	News report/blog Sports report Personal/diary blog Historical article Travel blog Short story Novel Biographical story/history Magazine article Obituary Memoir Other narrative	Description of a thing Informational blog Description of a person Research article Abstract FAQ about information Legal terms and conditions Course materials Encyclopedia article Technical report Other informational description/explanation	Opinion blog Review Religious blog/sermon Advice Letter to the editor Self-help Advertisement Other opinion	Description with intent to sell Persuasive article or essay Editorial Other informational persuasion	How-to Recipe Instructions FAQ about how-to Technical support Other how-to/ instructional	Song lyrics Poem Prayer Other lyrical	Discussion forum Question/answer forum Reader/viewer responses Other interactive discussion	Interview Transcript of video/audio Formal speech TV/movie script Other spoken
Quote from spoken language								
Title of a published work								

Table 3. Framework for categorizing main registers and sub-registers, adapted from Biber et al 2015a.

Table 3 also shows the structure of the framework and the slight adaptations that have been made for it to better suit the topic of this thesis. One differing factor from the original Biber et al. framework is the way reader comments are categorized as *reader/viewer responses* rather than making a separate note of their appearance in a comment section. This is because comments are so distinctively separate from the web register category of the web page they appear on. Since the focus of this study is on which specific registers the word forms appear in, it is more important to make note of the comment aspect of the web page than the category of the rest of the text. Additionally, a note is made if the word form appears in a title of some published work, instead of it appearing purposely as a part of a written text on a web page. Similarly, it is also noted if the form is used inside quote marks and therefore quotes spoken language. These additional observations will be discussed in the analysis if they occur notably in some web registers.

All the examples are entered into an Excel spreadsheet with links to their original web pages to reference their context. Each web page is visited manually and the register framework is followed to define which web register and sub-register the content can be categorized into. Drop-down lists for both categorizations are used in Excel for efficiency. The contents on the sub-register drop-down list are dependent on which category is chosen from the first list, as Biber et al. have noted that this makes the categorization into sub-registers more reliable (2015b: 1822). There is also space to mark if the specific studied expression appears inside a quote, or in a title of a song, book, or any other published piece of work. Instances that cannot be analyzed are disregarded and replaced by additional examples. These include examples where the context or meaning of the expression is unattainable, or the web page is inaccessible. As a result, the same number of analyzable examples are available of all expressions.

Dividing the examples into the different web register categories accounts for the extra-linguistic factors that surround the data. According to Berglund, if certain expressions are used interchangeably as synonyms, the expressions would be distributed evenly across the data and there

would be no distinctive differences in their usage patterns across text categories (2003: 69). The same principle can be applied here, as different usage patterns that may appear between *wanna* and *want to* and *gonna* and *going to* in the data would indicate differences in their use across different web registers. In contrast, similarities between the forms would indicate that there is no variation in the usage of these full and contracted expressions in online contexts. Comparing the contracted expressions to their full counterparts is therefore necessary to discover patterns in their use. This comparison will also make it possible to define whether the contractions or the full forms are preferred in each web register.

4. Results

Following the framework of Biber et al. (2015) the word forms were divided into the web registers and sub-registers one by one. This chapter will explain the results that followed from this categorization and demonstrate how the word forms are distributed across the different web registers and sub-registers. Results of each word form will be provided in this chapter, including both the contractions and full forms. This thesis focuses on the use of the contractions, as it aims to find out how their use is distributed across the web registers. Still, the results of the full forms are analyzed as well, so that their occurrences can be compared to those of the contractions.

Overall, the results indicate that all word forms occurred in the *interactive discussion* web register the most. Still, the uses of the contracted forms *wanna* and *gonna* appeared to be largely limited only to this category, though there were instances in the *narrative, lyrical* and *spoken* registers as well. The full forms were distributed more evenly across all the registers, despite *interactive discussion* being their most frequent register as well. A more detailed explanation of the distribution of each word form will follow in the subsequent sections. Each section will discuss the frequency of the word forms in the main web registers, as well as in their sub-registers. The tables and figures included here demonstrate the distribution of each word form individually. Chapter 5 will compare the word forms further and provide analysis on their usage patterns, as well as discuss the reasoning behind them. Hybrid registers will also be discussed in more detail there, although some information about their occurrences is presented in this chapter as well.

4.1. *Wanna*

Interactive discussion was largely the most common web register that used *wanna,* as Figure 1 indicates. This register had 141 occurrences of the word form, while the next register was *lyrical* with

less than a third of this number. *Spoken* and *narrative* registers also had a similar number of instances of the contraction, while the rest of the categories were even less common.

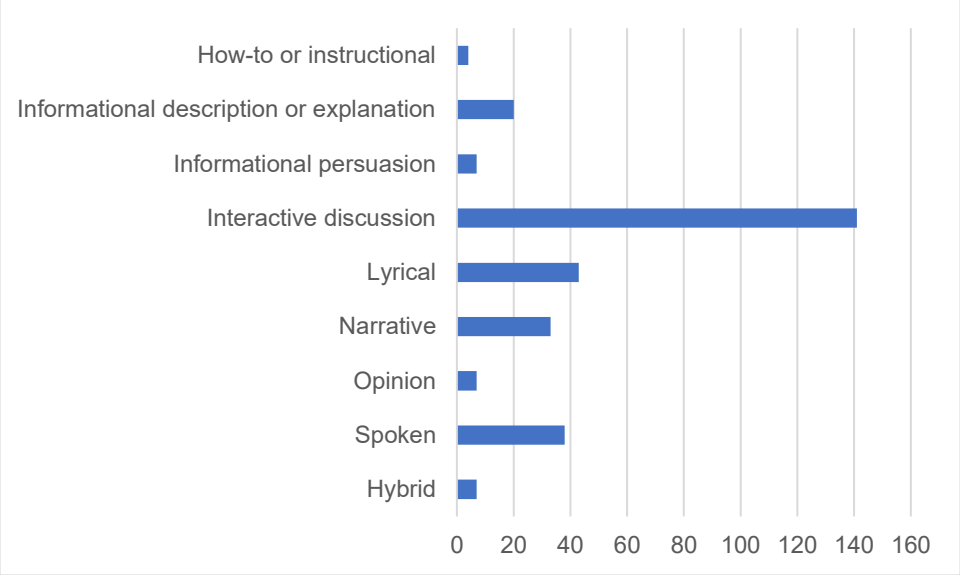

Figure 1. Distribution of *wanna* in the main registers.

Regarding the distribution across the sub-registers, the *interactive discussion* instances were divided between *discussion forum* and *reader/viewer responses*. In this thesis, comments were categorized as the latter, and their portion of the data was therefore 37 instances. *Wanna* occurred on discussion forums 99 times, which is a third of all the data. Lyrical texts were also well represented in the *wanna* data, with 41 instances appearing in song lyrics. Some examples of the use of *wanna* in its most common sub-registers are presented in Table 4.

Discussion forum	ppl just dont **wanna** hear that you have to diet strictly and train hard and maybe do bit cardio, they so lazy they dont even **wanna** HEAR how it is, let alone do it themselves.
Reader/viewer responses	I just **wanna** shout out a huge ThANKs for such inspiring and fun videos!
Song lyrics	I **wanna** love you but I better not touch, I **wanna** hold you but my senses tell me to stop.
TV/movie script	THELMA: I **wanna** get a job. I **wanna** work at Club Med.
News report/blog	“We **wanna** be the first business partner that a young entrepreneur has,” he said, noting that the DNA of a company is often set in the first 30, 60, or 90 days.
Personal/diary blog	I didn't **wanna** just compost it so I thought "Hey! I could make carrot cupcakes!"

Table 4. Examples of *wanna* in the data.

Most of the spoken instances appeared in the *TV/movie script* sub-register with 21 occurrences, but there were some instances of interviews and transcripts as well. The second largest *narrative* sub-register was *news report/blog*, but it should be noted that all ten of these instances appeared either inside quotation marks or in a title of a work. The contracted form was therefore never intentionally used within a news text itself. The other *narrative* sub-register that used *wanna* more than a few times was *personal/diary blog*. Nevertheless, there were only 12 instances of the contraction in this register, and its portion of the overall data is not particularly high.

Web register	**Sub-register**	**Occurrences**
How-to or instructional	Instructions	4
Informational description or explanation	Description of a person	4
	Description of a thing	11
	Informational blog	5
Informational persuasion	Description with intent to sell	4
	Persuasive article or essay	2
	Other informational persuasion	1
Interactive discussion	Discussion forum	99
	Question/answer forum	5
	Reader/viewer responses	37
Lyrical	Poem	2
	Song lyrics	41
Narrative	Magazine article	4
	News report/blog	10
	Personal/diary blog	12
	Short story	1
	Sports report	2
	Other narrative	4
Opinion	Advice	1
	Opinion blog	1
	Religious blog/sermon	1
	Review	3
	Other opinion	1
Spoken	Interview	9
	Transcript of video/audio	7
	TV/movie script	21
	Other spoken	1
Hybrid	Discussion forum/Other narrative	2
	Description of a thing/Opinion blog	3
	Discussion forum/Short story	1
	Opinion blog/Other narrative	1

Table 5. Distribution of *wanna* in the sub-registers.

Overall, the contraction was not overtly popular in any of the sub-registers except the ones that fall under *interactive discussion* and *lyrical*. As Table 5 indicates, it was used only in a few instructional texts, and its use was also scarce under the main registers *informational persuasion* and

opinion. There were also a few instances of *wanna* in the *hybrid* register. Some of these hybrid combinations occurred more than once, but overall the hybrid portion of the data was small. These occurrences will be discussed in more detail in section 5.3.

4.2. *Want to*

The distribution of the full form *want to* was more balanced across the main registers. Figure 2 illustrates the variation in its use, with *interactive discussion* being the most common register with 78 occurrences. The register of *informational description or explanation* closely followed with 52 instances, while *opinion, informational persuasion* and *how-to or instructional* all had around 40 occurrences. *Want to* was also used in the *narrative* web register 31 times, while it was less frequent in the rest of the categories.

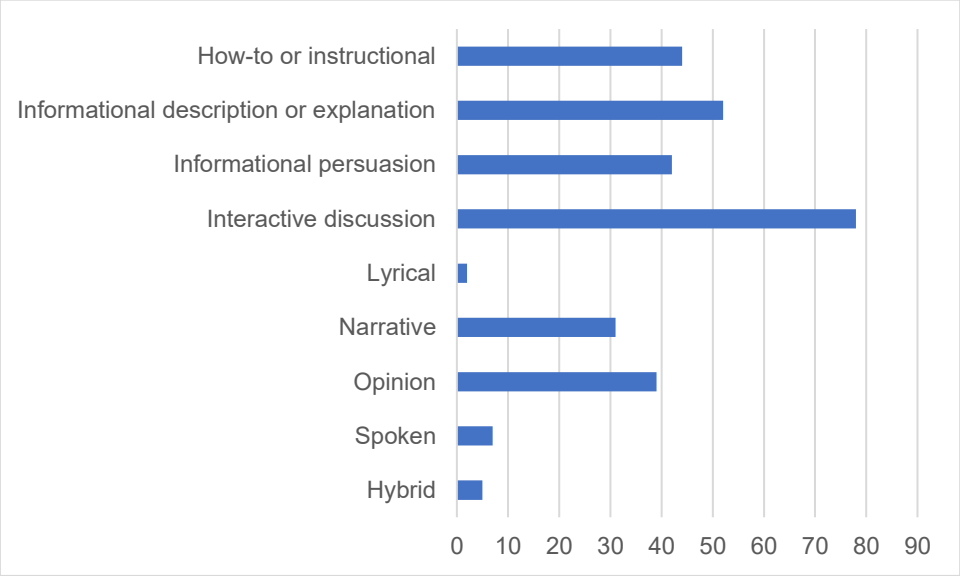

Figure 2. Distribution of *want to* in the main registers.

Focusing on the sub-registers, Table 6 shows that the distribution under the *interactive discussion* category was divided between *discussion forum* with 51, and *reader/viewer responses* with 20 instances.

Web register	**Sub-register**	**Occurrences**
How-to or instructional	How-to	12
	Instructions	32
Informational description or explanation	FAQ about information	4
	Description of a thing	22
	Informational blog	22
	Legal terms and conditions	2
	Research article	1
	Other informational description/explanation	1
Informational persuasion	Description with intent to sell	14
	Persuasive article or essay	19
	Other informational persuasion	9
Interactive discussion	Discussion forum	51
	Question/answer forum	7
	Reader/viewer responses	20
Lyrical	Poem	2
Narrative	Biographical story/history	1
	Magazine article	4
	News report/blog	8
	Novel	1
	Personal/diary blog	8
	Short story	2
	Travel blog	1
	Other narrative	6
Opinion	Advertisement	1
	Advice	9
	Opinion blog	4
	Religious blog/sermon	2
	Review	16
	Self-help	1
	Other opinion	6
Spoken	Interview	3
	Transcript of video/audio	4
Hybrid	Other narrative/Description of a thing	1
	Personal/diary blog/Instructions	1
	Other narrative/Interview	1
	Instructions/Other informational description/explanation	1
	Other narrative/Review	1

Table 6. Distribution of *want to* in the sub-registers.

The most common sub-registers of *informational description or explanation* were *description of a thing* and *informational blog*, with *want to* occurring in both 22 times. The full form was also used in *instructions* 32 times, and in *how-to* texts 12 times. Texts under the *informational persuasion* register were quite evenly divided, as *description with intent to sell* had 14 instances and *persuasive article*

or essay had 19. *Want to* was also quite equally distributed across the *narrative* sub-registers. Almost the same can be said of the *opinion* sub-registers, with only *review* standing out with 16 instances. Examples of the use of *want to* across its most frequent sub-registers are presented in Table 7 below.

Instructions	Once you have decided what you **want to** do, then start the program.
Description of a thing	If you **want to** know about the ins and outs of a letter of authorization, then you have come to the right place.
Informational blog	Keep in mind that if you **want to** use the ISOFIX anchorage points, you will need an ISOFIX compatible car seat.
Description with intent to sell	This modern wood desk features a minimalist charm that transforms your office into a warm and inviting space that you'll **want to** come back to day after day.
Persuasive article or essay	Most prospective clients **want to** say YES to getting help!
Discussion forum	There are a very large number of users who just **want to** take a photo of their dog, cat, family member and make them say and do "Crazy" stuff.
Reader/viewer responses	If you haven't maintained the landscape you may **want to** hire a property clean up service.
Review	This book made me **want to** throw it down in anger on how boring it was!!

Table 7. Examples of *want to* in the data.

Out of the main registers, *want to* was least frequently used in lyrical texts. It was also rarely used in the *spoken* register, and it only occurred once in all the different *hybrid* sub-registers. *Want to* also appeared rarely in some of the *narrative, opinion*, and *informational description or explanation* sub-registers, even though these main registers were well presented overall. Some of the least common sub-registers in these categories were *travel blog, advertisement*, and *research article*.

4.3. *Gonna*

The other contraction *gonna* occurred 124 times in the *interactive discussion* register, with *spoken* and *narrative* being the other categories with a few dozen instances. All other registers had less than 20 occurrences, as Figure 3 illustrates. It is notable that 31 of the 51 *narrative* instances appeared within quotation marks, indicating that over half of them were not used in the text of the web page itself but instead as a quote from someone else, possibly transcribed directly from spoken language.

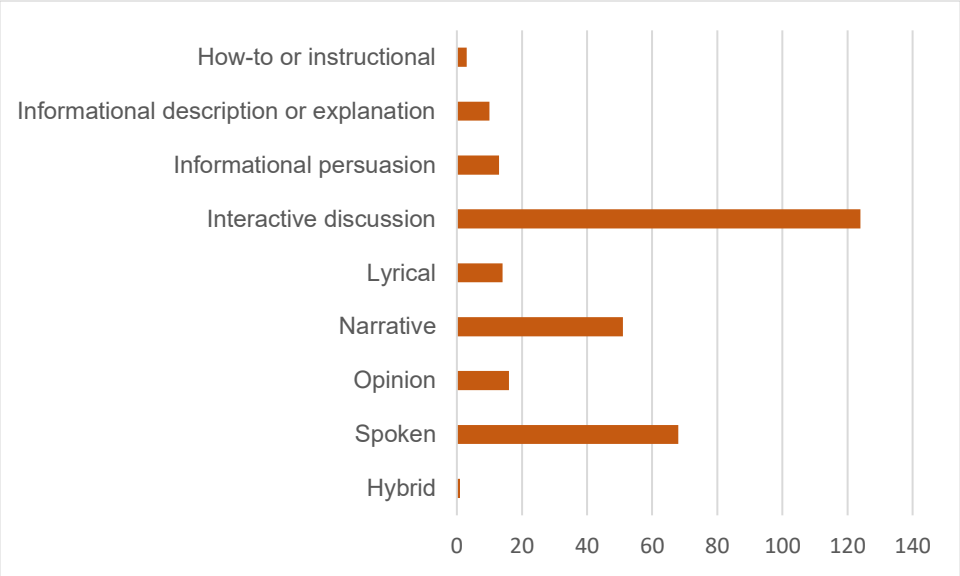

Figure 3. Distribution of *gonna* in the main registers.

Two of the *interactive discussion* sub-registers were well-presented in the data. *Discussion forum* was the most common sub-register with 78 instances, while *gonna* also occurred in *reader/viewer responses* 44 times. The contraction was also used in a few sub-registers under the *spoken* category, with the three most popular categories being *TV/movie script*, *transcript of video/audio*, and *interview*. Out of all the *narrative* sub-registers, the most common one was *news report/blog*. However, these instances occurred inside quotation marks, which implies that *gonna* was not used in news texts at all outside of quotes. Examples of the use of *gonna* in its most frequent sub-registers are visible in Table 8.

Discussion forum	Any help would be appreciated, cause my wife is **gonna** kill me if I don't get the field mowed!!!
Reader/viewer responses	I'm **gonna** keep it real, I don't like ANYTHING about wkend.
News report/blog	“I’m not **gonna** lie, it’s really stressful,” says Zhou, who’s led a few charity projects but nothing on this scale.
Transcript of video/audio	And you know, if I say that, I'm **gonna** be, you know, ostracized.
TV/movie script	REILLY: No-one in the park is **gonna** be able to see it from there.

Table 8. Examples of *gonna* in the data.

Table 9 shows that *gonna* was not commonly used in any of the *how-to or instructional* sub-registers, and the same can be said of the *opinion* sub-registers. The contraction also had only a few occurrences across the *informational description or explanation* sub-registers. Furthermore, categories under the *informational persuasion* register did not prefer to use the contraction either. It should also be noted that blogs are part of quite a few of the sub-register categories, but *gonna* was not particularly popular in any of them.

Web register	**Sub-register**	**Occurrences**
How-to or instructional	How-to	1
	Instructions	2
Informational description or explanation	Description of a person	2
	Description of a thing	5
	Informational blog	3
Informational persuasion	Description with intent to sell	5
	Editorial	1
	Persuasive article or essay	6
	Other informational persuasion	1
Interactive discussion	Discussion forum	78
	Question/answer forum	1
	Reader/viewer responses	44
	Other interactive discussion	1
Lyrical	Song lyrics	14
Narrative	Biographical story/history	2
	Magazine article	5
	News report/blog	17
	Novel	2
	Personal/diary blog	8
	Short story	3
	Sports report	4
	Other narrative	10
Opinion	Advice	2
	Opinion blog	8
	Review	2
	Self-help	1
	Other opinion	3
Spoken	Formal speech	1
	Interview	13
	Transcript of video/audio	21
	TV/movie script	31
	Other spoken	2
Hybrid	Other narrative/Description of a person	1

Table 9. Distribution of *gonna* in the sub-registers.

4.4. *Going to*

Interactive discussion was the largest web register that used *going to*, with 98 occurrences. Figure 4 indicates that other categories had less instances of this word form. There were no instances of *going to* in the *lyrical* register, and only three occurrences that were classified as *hybrid*.

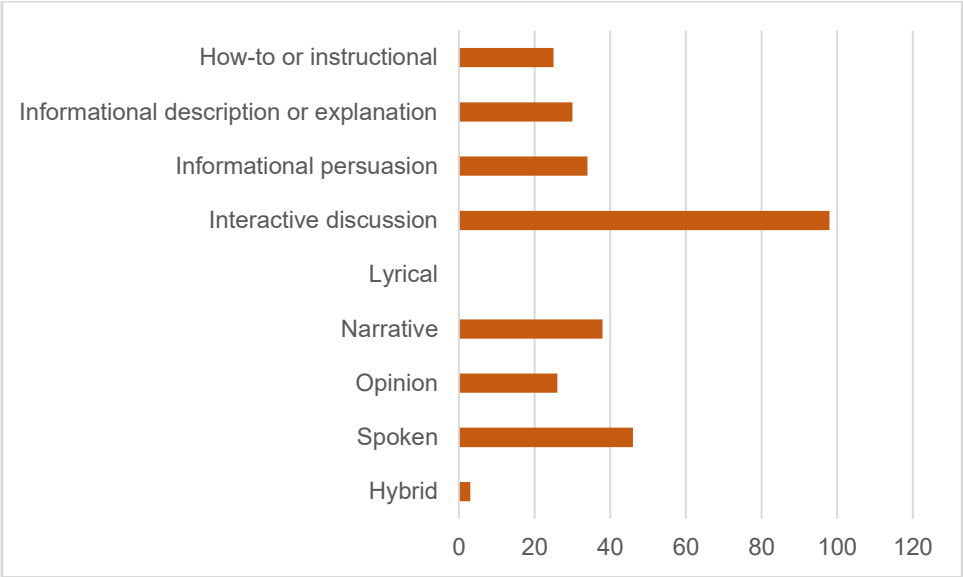

Figure 4. Distribution of *going to* in the main registers.

Nevertheless, there are similarities in the numbers between the rest of the registers, as the number of their instances were in the 20-40 range. This indicates a steady distribution of the word form across these web registers, although *interactive discussion* was still the most popular register.

How-to	Firstly, you're **going to** want to set your camera to manual mode.
Instructions	Keep this circuit mega-INTENSE; otherwise, you are not **going to** benefit from metabolism-boosting effects of HIIT.
Informational blog	In fact, because weeds are so absorbent, they are **going to** benefit just as much from your seasonal lawn care routine as your grass.
Persuasive article or essay	However, what they must realize is that the markets are not **going to** stabilize anytime soon.
Discussion forum	Are you **going to** paint 3 walls brown with an accent wall?
Reader/viewer responses	I am **going to** alter the humidity like you said and see what happens.
Personal/diary blog	I'm pretty sure I'm **going to** continue tracking calories after I'm finished, and probably will continue not to eat before noon (or after 8).

Table 10. Examples of *going to* in the data.

Out of the *interactive discussion* sub-registers, *going to* occurred on a *discussion forum* 74 times and in *reader/viewer responses* 21 times. The most common *spoken* sub-register was *transcript of video/audio*, where the word form was used 30 times. The distribution was quite even within the *opinion* and *narrative* sub-registers, with *personal/diary blog* having the most occurrences with 12 instances. *How-to* and *instructions* sub-registers also both had 12 instances. The most common sub-register of *informational persuasion* was *persuasive article or essay*, which had 24 instances. Some examples of the use of *going to* in its most frequent sub-registers are visible in Table 10.

Web register	**Sub-register**	**Occurrences**
How-to or instructional	How-to	12
	Instructions	12
	Recipe	1
Informational description or explanation	Description of a thing	10
	Informational blog	17
	Encyclopedia article	1
	FAQ about information	2
Informational persuasion	Description with intent to sell	7
	Editorial	1
	Persuasive article or essay	24
	Other informational persuasion	2
Interactive discussion	Discussion forum	74
	Question/answer forum	2
	Reader/viewer responses	21
	Other interactive discussion	1
Narrative	Biographical story/history	1
	Historical article	1
	News report/blog	8
	Novel	1
	Personal/diary blog	12
	Short story	1
	Sports report	5
	Travel blog	2
	Other narrative	7
Opinion	Advice	2
	Opinion blog	7
	Religious blog/sermon	2
	Review	9
	Self-help	1
	Other opinion	5
Spoken	Formal speech	1
	Interview	11
	Transcript of video/audio	30
	TV/movie script	2
	Other spoken	2
Hybrid	News report/Opinion blog	1
	Sports report/Opinion blog	1
	Advice/Description of a thing	1

Table 11. Distribution of *going to* in the sub-registers.

As Table 11 illustrates, *going to* did not occur at all in lyrical texts and for example the *TV/movie script* sub-register only had two instances of it. Several of the other sub-registers also used the word form only a few times, for example some of the *narrative* categories. Still, *going to* was well distributed across a variety of web registers, with *interactive discussion* having the most instances. However, the other registers had multiple occurrences of the word form as well, resulting in its overall distribution being quite even across the categories.

5. Discussion

The results presented in the previous sections will be discussed in more detail here. Overall, the findings appear to indicate that *wanna* and *gonna* are most common in informal contexts. Some language users lean towards the contractions in certain web registers, while others prefer to use the full word forms. The first two parts of the discussion section will compare the contractions to their full forms and outline the patterns in their use. Next the hybrid categories will be discussed in detail, taking into consideration their low frequency in the data. Following this, the contractions themselves will be compared with each other to see how similar their use is in different web registers, as well as to study possible variety in their use. This part will also discuss the defining characteristics of the registers where the contractions were the most frequent, and the relationship of spoken and written language in these contexts. Lastly, these findings will be examined in relation to internet language as a whole. Some of the limitations of the study will also be discussed, as well as the possible ways the study could be developed further to gain a deeper understanding of how spoken language is connected to written language online.

5.1. Comparisons of *wanna* and *want to*

Comparing the usage of the contraction *wanna* to its full form *want to*, it is apparent that *wanna* is more frequent in *discussion forum* and *reader/viewer responses*, making it overall the more popular form in the *interactive discussion* register. Still, both forms can be said to be prominent in this web register, as *want to* is quite frequent as well. However, the forms differ in figures across the sub-registers. The number of times *want to* is used in the *discussion forum* sub-register is half of the *wanna* figure. The same applies to *reader/viewer responses* (i.e. reader comments), where the number of *want to* occurrences is again half of the *wanna* instances. Furthermore, third of all the *wanna* occurrences were on discussion forums. The use of the contraction is therefore largely restricted to

interactive discussions, although it is used to some extent in other registers as well. The full form is used overall notably less than *wanna* across all the *interactive discussion* sub-registers. Interactivity and the number of participants therefore have significant roles when defining the online contexts where *wanna* is used the most.

Wanna was also more frequent in the *lyrical* and *spoken* registers, while the *want to* occurrences are relatively low in these registers. The spoken language contraction was expected to occur in these specific registers, as they are strongly linked to speech. Despite this connection to speech, the two registers differ distinctively. The mode of lyrical texts is originally written, while the *spoken* register has been originally produced in the spoken mode. Still, both are meant to be spoken out loud in one way or another.

The contents of the lyrical texts are expressed vocally after being written down, for example in the sub-register *song lyrics*. Song lyrics consist of text that is meant to be sung out loud, and in this data these lyrical texts have been later published on websites that collect a variety of popular song lyrics in one place. These websites are not necessarily the first place where these types of texts have been published, but instead a platform where the lyrical texts are distributed for the enjoyment of the internet users.

The *spoken* register has this in common with the lyrical texts, as their origin is also typically outside the internet and they are only distributed to a different audience through this medium. Some of the *spoken* sub-registers have been originally spoken out loud and then written into a text for the internet, for example the *transcript of video/audio* sub-register. These texts are often directly transcribed from spoken language and then converted into a written form. Other spoken sub-registers behave in an opposite way, as is the case with *TV/movie script.* This sub-register has been originally written as a text meant to be spoken out loud by actors. In this data, its original form has also been distributed on the internet in a way similar to the lyrical texts. Out of all the *spoken* sub-

registers, *wanna* was used the most in this *TV/movie script* category. These occurrences of the contraction can therefore be explained by its relation to speech.

While the use of *wanna* was largely restricted to only a few registers and sub-registers, *want to* appeared more consistently across the categories. *Want to* was used often in the *interactive discussion* register, but this number was also closely followed by those of other registers. The registers that showed more usages of *want to* than *wanna* were *how-to or instructional*, *informational description or explanation, informational persuasion,* and *opinion*. The full form was therefore used more in descriptive texts that had an informational purpose and were more formal. The word form was also more common in instructional texts that often have an exact purpose of informing the reader on how to succeed in an activity. The full form is therefore more popular in texts with a specific communicative purpose, while the contraction appears in informal texts that do not have as defined meaning and are more clearly connected to conversation and speech.

Interestingly, *want to* was used more often than the contraction in reviews, under the *opinion* register. *Wanna* could have been expected to occur often in this context, as reviews can be compared to reader-viewer responses to some extent, although they do differ in the number of participants. Both are typically short texts written online as a reaction to something else, but perhaps the target of the reaction differs between these two sub-registers and also affects the purpose of the text. Reader-viewer-responses often react to content that exists online, such as a blog text or news report. In contrast, reviews can express the person's opinion about anything in the online or offline world. The topic of the review could therefore be a lawnmower that a person has purchased, or an online course about programming. Reviews are written online, but their topic is not strongly connected to just the internet. Their purpose is to express opinion, while reader comments do not have a defined purpose in the register framework.

5.2. Comparisons of *gonna* and *going to*

Similarly to the other contraction, *gonna* also has more occurrences in the *interactive discussion* register than its full form. Nevertheless, the difference here is not as notable as it was between *wanna* and *want to. Going to* is used more often in this register than the other full form, and the contraction is used slightly less. *Gonna* and *going to* also have almost identical numbers in the *discussion forum* sub-register. However, it is noteworthy that while a third of all the *going to* instances appear to occur in *interactive discussion*, it is still used less in this register than its contraction. Their frequencies also differ in the sub-registers, as *gonna* is used twice as often in *reader/viewer responses*. The contraction is therefore preferred across the *interactive discussion* sub-registers, although the difference is not as striking as it was between *wanna* and *want to.* The possible reasoning for this will be discussed in section 5.4. Nonetheless, interactivity and the number participants affect the distribution of this contraction as well. *Gonna* is used most often in situations where some sort of information is distributed across multiple participants in a casual setting, which is the most common communicative purpose of this contraction. The contraction's preferred situation of use is therefore interactive conversation. This differs from some of the other registers, where one person or author is performing through text to an audience that is not expected to communicate directly back.

Gonna was more popular than its full form in only a few registers besides *interactive discussion*. The contraction occurred more often in *TV/movie script* sub-register and overall in other *spoken* sub-registers, as was also noted about *wanna*. The contraction was also used in the *lyrical* web register multiple times, while *going to* did not have a single occurrence in this register. As was mentioned in the previous chapter, the contraction here also appears to be used often in registers that either originated directly from spoken language or aim for the written text to be spoken or sung out loud. This is especially notable in the lack of *going to* occurrences in the *lyrical* register.

While the use of *gonna* was restricted to only a few registers, its full form was used across a variety of registers. *Going to* was used more often than its contraction in the *how-to or*

instructional register, as well as in *informational description or explanation, informational persuasion,* and *opinion.* These patterns are similar to those of *want to,* which indicates that the two full forms are both used quite consistently across different descriptive and informational registers.

5.3. Hybrids

All the word forms had occurrences that could not be divided strictly into just one category. These instances had the characteristics of multiple web registers, which defined their specific hybrid register. It was not possible to know beforehand which combinations of registers would be the most common in this data, and to what extent. As a result, these register pairs were not listed among the others in the web register framework. Their categorization was dependent on the data itself, and it was not as fluid as that of the other examples. Chapter 4 illustrated the occurrences of each word form across all the web registers, and the tables specified how often each example was used in a hybrid register. As the figures demonstrated, *hybrids* were overall not particularly frequent in the data.

The hybrid registers had seven instances of *wanna* and five of *want to*, while *gonna* occurred only once and *going to* three times. These numbers are very low relative to the 300 examples of each word form. It is possible that the data was quite straightforward to categorize, which may have resulted in the examples fitting easily into the pre-existing registers. This could be the case especially with the *interactive discussion* register, which had overall the most instances in the data. This register and its sub-registers can be clearly categorized due to their distinctive characteristics. Other registers may not be as easily characterized based on their communicative purposes, and their categorization can result in them being defined as *hybrids* (Biber et al. 2015a: 40).

According to Biber et al, *Narrative + Informational description/explanation* and *Narrative + Opinion* are the most frequent hybrid pairs in the searchable web, with *Informational description/explanation + Opinion* also occurring often (2015a: 33). In this study, all three of these

hybrids occurred the most out of all combinations. These results of the study therefore conform to the findings of Biber et al. to some extent. However, no actual conclusions can be made about how they are connected to the use of the contractions, as their overall numbers are so low in the data.

5.4. Patterns and variation

The most notable pattern in the data is that the contractions *wanna* and *gonna* behave quite similarly across all the registers. They are used mostly in the *interactive discussion* and *spoken* registers, and not very often in other parts of the internet. Figure 5 illustrates the exact percentages of each word form in the web registers. However, the *narrative* portion of this figure is slightly misleading, as most of these occurrences of the contractions occurred within quotation marks and not directly in the written text of this register.

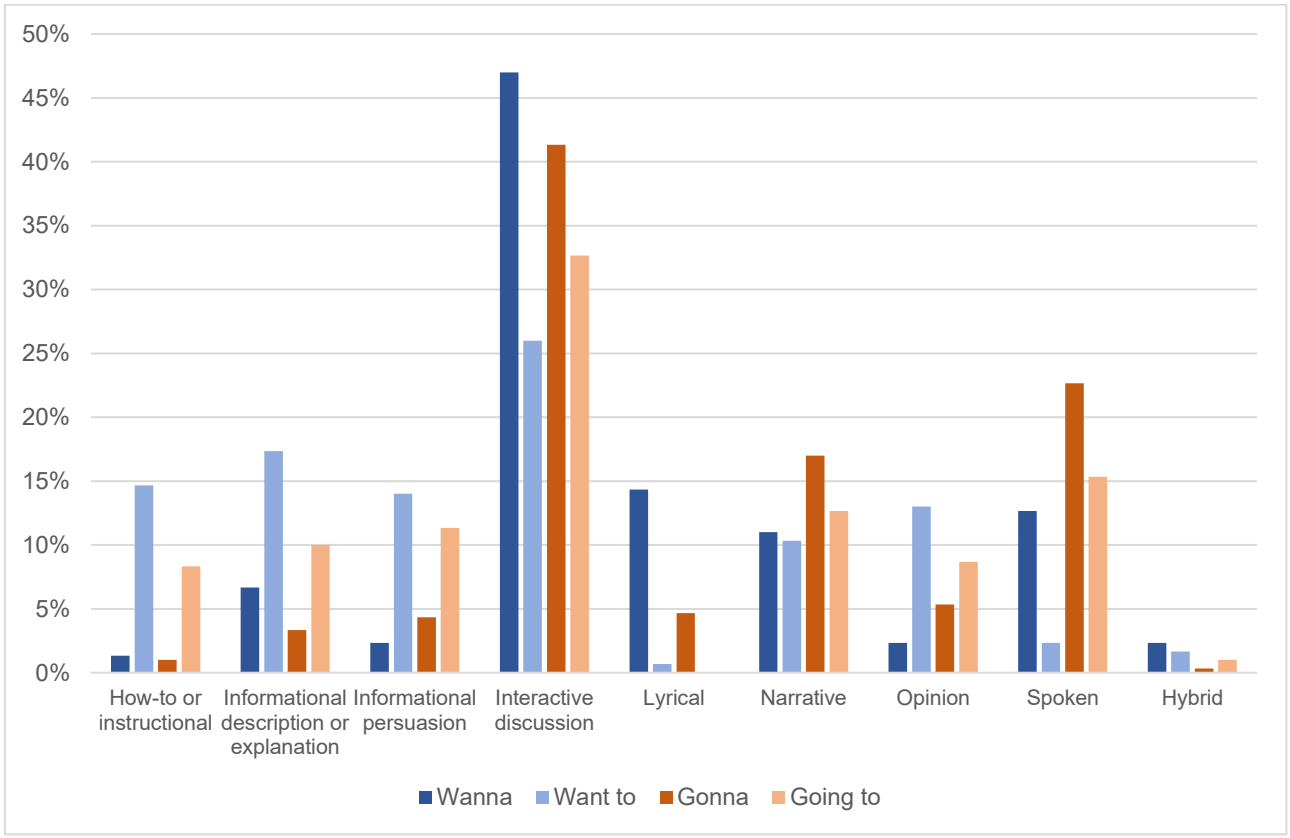

Figure 5. Percentages of the word forms in the main registers.

Majority of the language users in the *interactive discussion* register prefer to use the contractions rather than the full forms, in both *discussion forum* and *reader/viewer responses*. Still, the difference here between *gonna* and *going to* is not as distinctive as it is between *wanna* and *want to*. Based on these numbers, one might argue that *gonna* is a more formal contraction than *wanna*. However, perhaps the more likely explanation for this is that *going to* is used generally in more informal contexts than *want to*. Nonetheless, the contractions are still used more than their full forms in *interactive discussion,* despite some differences in the figures.

Interactivity and the number of participants play an important role in this, as *interactive discussion* is the most popular register of both contractions. This form of communication involves multiple participants, and there is a flow of information that takes place between them. As a result, the participants influence one another and the communication is not one-sided. The other registers have either a single author or co-authors, but there is no sense of interactivity in those contexts. *Interactive discussion* therefore differs distinctively from the other web registers with its conversational nature. The conversation that takes place in this setting differs from an offline conversation in a variety of ways, and these differences will be described next to give an idea about how spoken language is connected to this register. This will illustrate how the *interactive discussion* register resembles spoken language, but also how it aims to fill the communication gaps of a written conversation.

Firstly, there are multiple participants that take part in the discussion in an online setting, and it is not tied to any time or place. There can be restrictions on who can participate in the discussion, but this is not always the case. Typically, the only restriction is the requirement for the person to have a registered account on the website that enables their participation, for example on a discussion forum. This can also apply to *reader/viewer responses*, although some websites also allow anonymous comments that do not require any registration. Nevertheless, these restrictions on who

can take part in the interactive discussion are not particularly exclusive and they allow almost anyone to participate. In some cases, people may be excluded from the conversation because they do not belong into a specific social group and as a result, they do not have entry to the discussion platform. However, interactive discussions are more often openly accessible.

These conversions do also have different rules to those of a face-to-face discussion and there are challenges when it comes to distinguishing when the conversation will end and if all participants have been acknowledged in the discussion. As noted previously, Herring (1999: 10) has explained that this type of communication must compensate for the lack of clear turns between the participants, as well as for the shortage of feedback. Discussion forums and comments can distinguish between the turns in a way, as each message will appear in a specific place in the discussion. The discussion in this register mimics conversation in the way the participants respond to the topic on their own turn.

1)	if ya **wanna** find some more great blues men...your **gonna** have to dig, dig...get to the source...once you find that...the rest will follow, and you'll find some truly talented bluesmen...
2)	Someone help me! lol I really **wanna** understand how people do it because I get ideas for mixes everyday but I never have the acapellas to make them so this sort of knowledge will work miracles!
3)	If it's your first kit you're probably not **gonna** notice any noticeable differences, but you're **gonna** have a hell of a time putting the thing together compared to buying a legitimate kit.
4)	Guess what guys, you're skin's probably **gonna** look like shit the first few days, but after week one, you'll be like, "WHOAAAAA, is that me?!"

Table 12. Examples of *wanna* and *gonna* on discussion forums.

Discussion forums have separate messages from the participants in a chronological order, and there can be multiple messages from the same users. Table 12 illustrates some uses of the contractions *wanna* and *gonna* in this discussion forum setting. These messages resemble the turns of a spoken conversation, although the discussion may not be as fast paced. Similarly, reader comments occur typically also in a chronological order in their own specific section of the web page. There can be multiple comments from the same users in this context as well, and there are turns that are defined by the separate comments. Still, the conversation in this sub-register is perhaps more disconnected from

the other participants, as the responses are usually aimed towards the topic on the main web page rather than the other comments. Some examples of the use of the contractions in these types of comments are presented in Table 13 below.

5)	If you **wanna** talk bad about a country and its people, at least know something about it.
6)	I just **wanna** say how much I enjoy these videos and thank you guys for another year of fun.
7)	In a few days I'm travelling to NYC with my toddler.**gonna** be very useful!Love, love, love your blog!!!
8)	I love this tutorial. My husband has a cork board that he never uses and I'm **gonna** try and snag it out of his work area for a felt board 😉

Table 13. Examples of *wanna* and *gonna* in reader comments.

Occasionally there are also conversations that only concern the commenters. However, a participant may not consider what all the previous messages have discussed and could only respond to one specific topic or message in the discussion. As a result, certain turns can go unnoticed in the discussion as later messages ignore them and move on to something else. This is also linked to the shortage of feedback that was mentioned by Herring (1999).

Some of the participants do receive feedback in some way, but this does not apply to all the people who take part in the conversation. There is no pressure to reply to everyone in an online discussion with multiple participants, as there is no face-to-face contact taking place. The ignored person may not participate in this specific conversation again because they could have left the message board already, unlike in a spoken conversation where they would most likely still be in the same physical location. Again, this is linked to the lack of temporal and spatial restrictions in interactive discussions.

As the conversation does not have the typical restrictions of a face-to-face conversation, the participants can also take part in multiple conversations at the same time, as Herring (1999: 10) has stated. The conversations can continue even after a long time has passed, either by the same participants or someone else completely. One example is a discussion forum thread that has quieted down over time, but that a new user discovers and as a result the conversation resurfaces and become

active again. Typically, discussion forums save the textual records of older discussions, either in an archive or among the rest of the discussion threads. This can allow internet users to revisit older conversations later, although it is possible that they are closed off from further comments. Nonetheless, this means that there are no limitations as to how many conversations the person can take part in at the same time. Similarly, there is also usually no time limit to how long the conversation can last. It can last from a few minutes to multiple years, and the participants can vary during this time or stay the same. There is also no clear ending to the conversation, as the participants may return to the topic later. Thus, these interactive discussions do not have the same limitations as those of a face-to-face conversation, which gives their participants more freedom. Nevertheless, it can also result in restrictions, as was mentioned earlier.

Interactive discussions are also open-ended due to there being no guarantee that all the participants read or react to a message that is aimed towards them. The length of a delay between turns is not defined by the setting of a conversation in the same way it is in face-to-face conversations. In these conversations a long delay between turns indicates that the discussion is over, while no such rules exist in online conversations (Hård af Segerstad 2003: 30). As a result, online conversations have a sense of uncertainty.

The mode of the *interactive discussion* register also differs from that of a spoken conversation, as it is originally written, rather than spoken. The same also applies to the *lyrical* register, where the *wanna* contraction was quite frequent. Nonetheless, as mentioned earlier, the *lyrical* register is also linked to speech, especially in its most popular sub-register *song lyrics*. Another popular register among the contractions was *spoken*, which in turn originates in the spoken mode and is later distributed on the internet in a written form. The contractions were linked to speech in other registers as well. Another example is the *transcript* sub-register, where both contractions occurred in direct transcriptions of spoken language. Additionally, the contractions occurred often in scripts, which mainly consist of lines of texts that are to be said out loud by actors.

It is also notable that while *gonna* and *wanna* did occur within the sub-register of *news report/blog*, these instances actually took place within quotation marks. The contractions therefore occurred as part of reported speech, further strengthening their connection to spoken language. They were also used often in interviews, which report speech much in the same way as the *news report/blog* sub-register does. Berglund (2005: 84) has previously noted that *gonna* is the most frequent in contexts that resemble speech and include quoted speech within them. These findings are relevant to this study as well, concerning both contractions. Still, *interactive discussion* remains as their most popular web register.

The registers that used the contractions the most are therefore all linked to spoken language either directly or by imitating spoken conversation. This further defines the contractions *wanna* and *gonna* as features typically associated with spoken language. Consequently, they can also be defined as informal language. McCulloch (2019: 11) has mentioned that informal writing typically saves time by using shorter words and phrases and typing less letters. As a result, higher frequencies of contractions in certain web registers indicates that these situations are more informal than the ones that prefer the full forms. *Interactive discussion* had the most occurrences of the contractions, and therefore it can be clearly classified as an informal register.

Informal writing is typically discouraged in school but still used in notes and other pieces of text, as mentioned by Baron (2008: 46). Informal language is also strongly represented in certain parts of the internet, and it is a defining characteristic of the *interactive discussion* web register. The commonness of informal writing online is therefore an important feature of internet language, as it differs from the language norms offline. The participants in an online conversation do not necessarily aim to write English according to all the traditional rules, but instead to suit their own purpose. Consequently, this purpose is often to communicate with others by using all means possible. Sometimes this means borrowing features from spoken language and implementing them in the written text, or perhaps showing emotions by using emoticons in the same context. Spoken

conversations can be imitated in multiple ways, and one way to do this is by adapting text to look more like speech through informal language features.

Interestingly, the contractions were not particularly popular in any of the blog registers, though there were some instances of them. This low frequency of contraction occurrences is intriguing, as blogs are often viewed as informal platforms for expressing oneself that resemble public diaries to some extent. Perhaps blogs do enable conversational communication, but the language that is used is not particularly informal and does not favor contractions. On the other hand, the results did indicate that comment sections on different web pages used the contractions often. Thus, the person who writes the blog posts follows the rules of written texts and adopts the role of an author, while the commenters do not feel the need to do the same and instead they use informal word forms and contractions more in their replies. Blogs can therefore have two sides: on one side is the author that writes conversational texts but still follows the rules of formal writing. On the other side, the audience reacts to this text as if it had been spoken out loud, and they write reader comments that utilize spoken language features in a written mode. Consequently, Crystal (2011: 27) has noted that blogs solicit responses from the readers, but they are considered optional. Blogs typically pursue comments from the readers and in this way resemble interactive conversations, but they are not always successful and as a result this type of communication may be one-sided.

Interactivity is therefore a clear factor that defines whether an informal or more formal word form is chosen online. According to Yates (1996: 39), internet language harnesses information into a textual form that resembles writing more than speech, but can also use literate forms to fit this orally-oriented and social interaction. In other words, the participants of an interactive discussion adapt to communicate in a way that is suitable for the situation. Concerning the results of this study, it means using spoken language forms in a written text so that the discussion can more closely resemble an actual face-to-face conversation.

Studies concerning chat conversations have discovered that the dialogues in these settings aim to produce visual and auditory effects in a written form. The participants can appear almost maniac in the way they stretch language so that it can simulate speech (Werry 1996: 58). These strategies try to compensate the restrains of the communicative interaction, meaning the temporal and spatial restrictions among others (Werry 1996: 61). Chat conversations were not studied in this thesis, as they are a more private mode of communication and not included in the iWeb corpus that consists of public texts collected from the searchable web. However, they do share certain features with the *interactive discussion* register that was prevalent in the study. Discussion forums and comment sections are not as constrained, as their way of communication is less rapid than that of chat conversations. Still, they share the desire to communicate effectively in writing according to speech conventions. As mentioned by Werry, this can be done by producing the visual and auditory effects in a visual form, which means including contractions in a written text.

There is some variation within the web registers as well, as is visible for example in the *interactive discussion* register where both contractions and full forms were used, though in differing numbers. The contraction frequencies varied in the *lyrical* and *spoken* registers, although both contractions were prominent in these web texts. *Wanna* was more frequent in the *lyrical* register, while *gonna* is used more in the *spoken* register. As a result, it could be argued that *wanna* is more often used in expressive situations meant for entertainment, while *gonna* is used in spoken situations that aim to provide information rather than amuse others. This is also supported by the distribution of the contractions across the sub-registers. *Wanna* appeared in *song lyrics* 41 times, and its most popular *spoken* sub-register was *TV/movie script* with 21 occurrences. Consequently, these sub-registers typically aim to entertain the audience. In contrast, *gonna* was used in *song lyrics* only 14 times while its usage in the *spoken* sub-registers was divided between *TV/movie script, transcript of video/audio*, and *interview*. Out of these three, *TV/movie script* had the most occurrences, and as a result it can be argued that *gonna* is used in entertaining contexts as well. However, it also appears in transcripts and

interviews, which can be viewed as entertainment but also as informative situations. It is therefore apparent that *gonna* is used in slightly more informative spoken contexts. Nonetheless, all of the *lyrical* and *spoken* sub-registers that used the contractions the most are linked to speech in some way. As noted earlier, this connection to speech is important when defining the usage patterns of these word forms in contrast to their full forms.

The full forms *want to* and *going to* were quite consistently used across different web registers. This supports the hypothesis of this study that their use is more steadily divided across the registers, contrastively to the contractions. Out of the 300 examples, six of all the nine web registers used *going to* in the 20-40 occurrences range. Concerning *want to*, the same can be said of five registers. In addition to this, both word forms were used in *interactive discussion* 80-100 times. The only web register that had almost no occurrences of both forms was *lyrical.* The main difference in the use of the full forms was the popularity of *going to* in the *spoken* register, which mainly consisted of transcripts. In comparison, *want to* was very rarely used in this register.

When speculating the possible reasons for why the contractions where chosen over the full forms in certain web registers, the physical restrictions of this platform should be considered. People typically type text quickly on the internet, which can result in errors. Using shorter forms, such as contractions, results in fewer typing errors and misspellings as the words that are typed are shorter and have less chances of mistakes occurring (Varnhagen et al. 2010: 730). However, this is not an apparent reasoning for the decision to use the contractions over the full forms. One possibility is that the users feel that the writing setting is informal enough for them to pick *gonna* over *going to*. Another reason could be that the language users are not aware that they are using a contraction in a setting where it does not typically occur, namely in a written text. It is possible that they are not familiar enough with the norms of written texts to know that the word form they are using is a contraction typically preferred only in spoken language. Still, perhaps the most likely reasoning is that the language users are not consciously choosing to use the contraction over the full form. They

are writing a piece of text online, and they feel that the contraction fits into the text naturally and they do not stop to think about why they choose to use it rather than the full form. This could be because of the language norms that exist in this particular setting. The situation is informal enough for the person to use the contraction typical to spoken language unconsciously. Thus, the conclusion that can be drawn from this is that the informal and speech-like settings of certain web registers encourage the use of this type of language imperceptibly.

Higher frequencies of the contractions in the informal register of *interactive discussion* may also be due to the lack of editing that occurs on these websites. Discussion forums and comment sections do face outside moderation to some extent, but the people behind the actual texts do not edit their own content in the same way the writers behind for example news reports do. Some more formal articles may have included the occasional *wanna* or *gonna* in their texts to begin with, but these informal word forms are not very likely to be included in the final version of the texts that have been published. In contrast, discussion forums and comments are written by individuals and the norms of these websites may actually encourage the use of informal forms. Furthermore, the individuals are not necessarily as strict about their language use, as more formal writers would be. The main purpose of the people behind these informal web registers is to communicate their point to the other participants of the conversation, and in this case expressiveness may be preferred over correct and formal spelling.

Hård af Segerstad discovered in her 2003 dissertation that the way language is used in web chats and instant messages reflects the activity where it is used. These synchronous communication situations require active participation from the people who take part in the conversation. Thus, language is used similarly to speech in these informal interactions. Hård af Segerstad (2003: 268) speculated whether these norms could also be transferred to other areas of written interaction. As the results of this study indicate, the same norms are also present in interactive discussions that are asynchronous and do not require simultaneous communication from the

participants. Interactive discussions resemble speech as well, even though their communication is not as instantaneous as that of web chats.

As mentioned earlier, Crystal (2011: 29) has defined internet language as a form of communication that is guided by speech conventions. This means that language conventions typical to spoken language occur in internet language as well, and this study has discovered that they are especially common in the informal web register of *interactive discussion*. This web register can therefore be said to resemble spoken language more closely, while other registers follow the rules of written language and prefer to use the full forms. This thesis has approached the speech conventions from the point of view of contractions, which means that the findings are limited to this topic. However, the findings can be applied to other spoken language features as well, indicating that the registers where contractions were especially common could also adopt other speech conventions. Therefore, the high occurrences of *wanna* and *gonna* in *interactive discussion* indicates that the use of other contractions would also be popular in this context, and perhaps slang words and filler words may occur as well.

5.5. General discussion

Certain web registers had zero or very few occurrences of the word forms, which does not necessarily indicate that these types of web pages are not well represented online. Instead, it can be explained by the size of the data set. The scarcity of some registers in the data implies that these web registers did not use the word forms often enough to stand out within this data set. Additionally, some types of texts, i.e. obituaries and memoirs, might not be well presented on the internet as a whole. As a result, their portion of the results is also non-existent. In contrast, certain sub-registers are well presented online, i.e. research articles, but the full forms *want to* and *going to* do not fit into the preferred formal writing style of these texts and therefore they may not occur in these sub-registers at all. Consequently, the informal contractions are not used in these sub-registers either. Advertisements

were also barely visible in the data, even though they do form a significant portion of the internet and can be assumed to use informal language to some extent as well. Their lack of occurrences can be explained by the fact that advertisements are typically in picture form. As Biber et al. (2015a: 23) have noted, these types of web pages are not part of the searchable web due to their lack of text, and as a result they are also not included in the iWeb corpus.

It should be noted that there was a significant loss of data that took place during the analyzing process. Concerning all analyzed word forms, approximately third of each data set was inaccessible due to outdated URLs or web pages being inaccessible. This meant that the context for these examples could not be retrieved, and therefore their web register or sub-register could not be defined. These web pages had been changed after the creation of the corpus, and as a result they had to be replaced with additional examples. This may have affected the results of the study to some extent, although the substitutive data was collected in the same manner and at the same time as the disregarded data.

Web as a platform is dynamic and constantly changing, as mentioned by Biber et al. and therefore the links and content can change as well (2015a: 17). Biber et al. also noticed the loss of data in their own study of categorizing web registers. Between the identification of the web pages and downloading of the documents, eight percent of all their data had become inaccessible. This happened in seven months and only concerned the broken URLs of the web pages. In the case of this study, the corpus data had been gathered in 2017 and the examples studied in this thesis were downloaded from the corpus in November 2019. This is a longer time period than in the other study, which can explain the even larger amount of lost data. Furthermore, the content of each web page had to stay the same so that the specific web register where the word forms were used could be determined. As a result, even slight changes in the URL or page content of a web page may have led to its dismissal from the analysis.

Another limitation of the study was the amount of the analyzed data. As the categorization of each individual web page had to be done manually, the data size had to be limited to 300 examples per word form. Considering the 100 inaccessible web pages linked to each word form, approximately 400 web pages had to be visited manually to determine the web register of each specific word form in their context, or to dismiss the example as unanalyzable. This resulted in a total of 1600 visited web pages when analyzing all four word forms. While analyzing this number of examples manually was laborious, the frequency of each word form was not particularly high when divided across all the used web registers and sub-registers. Therefore, the results of this study are compromised to some extent due to the small amount of data. These limitations could be resolved in future research by analyzing a larger set of examples of internet texts, possibly by using automatic register identification that Biber et al. (2015a: 41) have been working on. Using these tools was not possible in the scale of this thesis, but they could aid forthcoming studies that require effective categorization of web registers.

As Biber et al. (2015a: 22) have previously noted, dividing web pages into the main web registers is more reliable than their division into sub-registers. This should be taken into consideration when looking at the distribution of the word forms across the different registers. However, this was strongly linked to the categorization of *hybrid* registers, which were not widely used across the data as most of the examples could be divided into just one web register and sub-register. Additionally, larger patterns in the use of the word forms are visible even when just the main web registers are considered. Further analysis based on the sub-registers was done as well, but the main results of the study do not change when looking at only the larger web registers.

Nonetheless, the results could be different if the analysis was done by multiple participants, as the categorization in this thesis is the result of just one person's work that is bound to affect the result to some extent, even when following the theoretical framework. Comparing the analysis of multiple persons would result in an average categorization of each example and more

reliable results. A follow-up study could therefore strengthen the results of this thesis, or perhaps extend the angle by examining additional features of spoken language in a similar manner online. For example, other contractions could be studied and as a result more specific information about their use in a larger scale could be revealed. Further research could also look into the age of the language users in the different web registers. According to Leech et al, contractions are more common in the speech of younger speakers (2009: 106). Many pieces of research have also stated that contractions are more common in spoken rather than written corpora (Berglund 2005, Leech et al. 2009). As a result, it can be speculated that the people who did use the contractions in written texts are of young age. The data did not actually have any information about the age of the participants, which is to be expected as it has been gathered from all over the internet where the age is not apparent. Nevertheless, it could be interesting to see if it could be proven that the people who do use contractions online are in fact mostly younger speakers.

Despite some of the limitations of the study, the results have revealed that occurrences of spoken language features in written online texts define some web registers more informal than others. The definition of internet language has typically involved discussion about how it has features from both spoken and written language. This study took this definition further, and examined how exactly these spoken language features, in this case contractions, are present across the different areas of internet. Dividing the use of language online into web registers was a systematic way to approach this topic. Comparing the use of contractions and their full forms across the web registers lead to findings about which parts of internet language use the spoken language features most frequently, and which do not. As a result, it was possible to define some web registers as best representing informal writing online. More specifically, conclusions were made about how internet language is connected to spoken and written language.

Still, internet language cannot be defined as just one way of using language, as there is variety in its different registers. This variety is defined by its surrounding situational characteristics

and its particular communicative purpose. Concerning the contractions typical to spoken language, interactive and speech-like web registers use these spoken contractions more often than others, in comparison to their full forms. The full forms are fairly evenly distributed across all the registers, which indicates that their use is not tied to informal communication situations resembling spoken conversations, as is the case with the contractions.

Internet language itself has been defined often as a combination of written and spoken language. This thesis defines this language environment further: interactive discussions and web registers directly linked to speech resemble spoken language the most. These registers clearly prefer to use the contractions over the full forms. Contrastively, the rest of the registers are closer to typical written language. Nonetheless, this division between features of spoken and written language is not black and white. Features of both spoken and written language are present in almost all web registers. As a result, comparing and further analyzing their usage allows us to make conclusions about how the different parts of internet language behave. Thus, in some web registers there is tendency to lean more towards spoken language conventions, and in others towards written language. But overall, there are hints of both all over internet language. Internet language is therefore a mixed medium, as Crystal has defined it (2011).

Previous studies on *gonna* have stated that it is almost exclusive to spoken corpora. As this study shows, this is not the case. This contraction, as well as *wanna*, occurs in written web registers as well. The difference in these results can be explained by the fact that previous studies that used corpora have focused on formal writing rather than informal texts. Therefore, this study differs from the previous ones in its usage of informal writing, as well as in the number of written instances of the contractions that could be retrieved from the corpus.

Perhaps this behavior of using spoken language features to express oneself fittingly to the situation could spread to other web registers as well over time, and as a result change people's attitudes towards what is appropriate in certain types of written language. These attitudes have already

changed to some extent, as can be seen in the frequencies of *wanna* and *gonna* in certain parts of the internet. Nonetheless, the distribution of the contractions is mainly restricted to a few web registers, while their usage is not as acceptable in others.

6. Conclusion

To conclude, this thesis studied the use of the contractions *wanna* and *gonna* in internet language, as well as their full forms *want to* and *going to*. The aim was to examine how common the contractions are in web registers in comparison to their full forms, and what patterns can be found in their use. The use of contractions is not typically encouraged in writing, but they are proven to be common in spoken language. Previous studies on internet language have found its use of language to be innovative and informal, which is why this thesis examined the distribution of informal contractions in the searchable web. The iWeb corpus proved to be a suitable source for this data, as it also focuses on this part of the internet. For the same reason, the theoretical framework of Biber et al. (2015a & 2015b) was used to systematically divide the examples into web registers and their sub-registers.

In the beginning of this thesis, two research questions were defined. The first question concerned occurrences of the contractions across the web registers and the reasons for these numbers. The findings of the study indicate that *wanna* and *gonna* are both used most often in informal online contexts that imitated conversation. They were the most frequent in the *interactive discussion* register, and its sub-registers of *discussion forum* and *reader/viewer responses,* the latter of which is also known as reader comments. These occurrences can be explained by the conversational nature of these sub-registers, where the communication is guided by speech conventions. Interactivity and multiple participants are also defining characteristics of these sub-registers. However, these conversations differ from the face-to-face ones as they have no temporal or spatial restrictions. Consequently, this may also result in some advantages in interactive discussions. The participants of an interactive discussion can take part in multiple conversations simultaneously, and the conversation can take place over a longer period of time. Perhaps this is another reason why the contractions were used in these contexts, as the participants may feel that they have to compensate for the lack of intimacy from face-to-face conversations by adopting word forms directly from these situations.

The second research question aimed to find out how the distribution of the contractions differed from their full forms *want to* and *going to* in the web registers. The findings of the thesis indicate that the full forms were distributed more evenly across the web registers than the contractions. The contractions occurred mainly in the *interactive discussion* register, as well as in registers that were directly related to speech, namely the *spoken* register. The high frequency of the contractions in the second register was not a surprise, as these word forms originate from speech. The contracted forms were therefore preferred over the full forms in clearly spoken contexts, as well as in the informal and conversation-like sub-registers of *discussion forum* and *reader/viewer responses*. Thus, these situations resemble spoken conversations more than others, and their language norms are unlike those of formal written texts. The textual content of discussion forums and reader comments is also not edited in the same way as the content of more formal websites is.

These findings indicate that word forms that have typically been considered to belong only to spoken language do in fact appear in certain written texts online as well. The contractions *gonna* and *wanna* were the most popular in *interactive discussion* register, which may not explicitly encourage the use of informal language, but its users do not criticize the use of the contractions either. The distribution of the contractions was therefore quite limited across the different registers of internet language, although their use may become more acceptable in other registers as well over time. The use of contractions could spread for example to blogs, which did already have some occurrences of them in this study. However, the contractions were more prominent in their comment sections, which could lead to them being used more in the actual blog texts as well. Language conventions can always change, and other registers may begin to follow speech conventions more and as a result allow and encourage the use of informal language.

While the use of the contractions was clearly limited to only a few registers in this study, they did occur at least a few times in all main registers. This indicates that the different parts of internet language can all resemble both spoken and written language to some extent. Overall, internet

language cannot be treated as a singularity, as different parts of it prefer to use language in a variety of different ways. This study has discovered that some parts of the internet prefer to use informal language forms, while others lean more towards traditional written norms. The study of internet language should therefore be approached on the conditions of its different registers, rather than by treating it as one entity.

References

Baron, Naomi S. 2008. *Always on: language in an online and mobile world*. Oxford: Oxford University Press.

Barton, David & Carmen Lee. 2013. *Language online: investigating digital texts and practices*. London: Routledge.

Berglund, Ylva. 2005. *Expressions of future in present-day English: a corpus-based approach*. PhD dissertation, Acta Universitatis Upsaliensis. http://uu.diva-portal.org/smash/record.jsf?pid=diva2%3A166402&dswid=-5564 [accessed 12 November, 2019].

Biber, Douglas, Jesse Egbert & Mark Davies. 2015a. Exploring the composition of the searchable web: a corpus-based taxonomy of web registers. *Corpora* 10(1), 11–45. doi:10.3366/cor.2015.0065.

Biber, Douglas, Jesse Egbert & Mark Davies. 2015b. Developing a bottom-up, user-based method of web register classification. *Journal of the Association for Information Science and Technology* 66(9), 1817–31. doi:10.1002/asi.23308.

Bolinger, Dwight. 1981. Consonance, Dissonance, and Grammaticality: The Case of *Wanna*. *Language & Communication* 3(1), 189-206.

Collot, Milena & Nancy Belmore. 1996. Electronic Language: A New Variety of English. In Herring (ed.), *Computer-mediated communication: linguistic, social and cross-cultural perspectives*, 13–28. Amsterdam: Benjamin.

Crystal, David. 2006. *Language and the Internet*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139164771.

Crystal, David. 2011. *Internet linguistics: a student guide*. London: Routledge. doi:10.4324/9780203830901.

Davies, Mark. 2017. *The iWeb corpus*. www.english-corpora.org/iweb. [accessed 11 November, 2019].

Hård af Segerstad, Ylva. 2003. *Use and adaptation of written language to the conditions of computer-mediated communication*. University of Gothenburg.

Herring, Susan C. 1996. *Computer-mediated communication: linguistic, social and cross-cultural perspectives*. Amsterdam: Benjamin.

Herring, Susan C. 1999. Interactional Coherence in CMC. *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences* 2, 1–13. doi:10.1109/HICSS.1999.772674.

Hudson, Richard. 2006. Wanna Revisited. *Language* 82(3), 604–27. doi:10.1353/lan.2006.0144.

Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: a grammatical study*. Cambridge: Cambridge University Press.

Lorenz, David. 2013. On-Going Change in English Modality: Emancipation Through Frequency. *Zeitschrift für Literaturwissenschaft und Linguistik* 169(1), 33–48. doi:10.1007/BF03379871.

Machová, Dagmar. 2015. The degree of grammaticalization of gotta, gonna, wanna and better: A corpus study. *Topics in Linguistics* 15(1). doi:10.2478/topling-2015-0005.

McCulloch, Gretchen. 2019. *Because internet: understanding the new rules of language*. New York: Riverhead Books.

Varnhagen, Connie K., Peggy McFall, Nicole Pugh, Lisa Routledge, Heather Sumida-MacDonald & Trudy E. Kwong. 2010. Lol: New language and spelling in instant messaging. *Reading and Writing* 23(6), 719–33. doi:10.1007/s11145-009-9181-y.

Verplaetse, Heidi. 2003. What you and I want: A functional approach to verb complementation of modal want to. In Facchinetti (ed.), *Modality in Contemporary English*, 151–90.

Werry, Christopher C. 1996. Linguistic and interactional features of internet relay chat. In Herring (ed.), *Computer-mediated communication: linguistic, social and cross-cultural perspectives*, 47–63.

Yates, Simeon J. 1996. Oral and Written Linguistic Aspects of Computer Conferencing: A Corpus Based Study. In Herring (ed.), *Computer-mediated communication: linguistic, social and cross-cultural perspectives*, 29–46.